
Metadades

PID_00271440

Blas Torregrosa García

Temps mínim de dedicació recomanat: 1 hora



**Blas Torregrosa García**

Enginyer en Informàtica i màster universitari en Seguretat de les Tecnologies de la Informació i de les Comunicacions (MISTIC) per la Universitat Oberta de Catalunya (UOC). Especialitzat en ciberseguretat. Professor col·laborador del màster de Ciència de Dades de la UOC i professor associat a la Universitat de Valladolid (UVA).

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats pel professor: Ferran Prats Carrasco (2020)

Primera edició: febrer 2020
© Blas Torregrosa García
Tots els drets reservats
© d'aquesta edició, FUOC, 2020
Av. Tibidabo, 39-43, 08035 Barcelona
Realització editorial: FUOC

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit dels titulars dels drets.

Índex

Introducció	5
1. Què són les metadades?	7
1.1. Diferència entre dades i metadades	8
1.2. Tipus i funcions	8
2. Estàndards de metadades	10
2.1. Tipus i estructura de metadades	10
2.2. Dublin Core Metadata Initiative	11
2.3. IEEE/LTSC LOM i IMS Global Learning Meta-Data	12
2.4. Més sobre estàndards	14
Bibliografia	15

Introducció

Les metadades van ser conegudes al món l'any 2013 quan Edward Snowden, un antic empleat de l'Agència Central d'Intel·ligència (CIA, per les seves sigles en anglès), va revelar que l'Agència Nacional de Seguretat (NSA, en anglès) estava duent a terme una vigilància massiva en les activitats de correu electrònic i dispositius mòbils dels seus ciutadans. La informació que estava registrant no era el contingut de les converses o missatges sinó les **metadades**, és a dir, informació sobre aquestes converses o missatges.

En aquest mòdul veurem què són les metadades, la forma com es tracten i alguns estàndards sobre les metadades.

1. Què són les metadades?

Una definició de metadades freqüentment usada és «**dades sobre dades**». Aquesta és una definició literal, encara que no excessivament aclaridora.

L'any 2004, l'Organització Nacional d'Estàndards d'Informació (NISO, per les seves sigles en anglès) va definir les metadades com a «informació estructurada que descriu, explica, localitza o fa més fàcil recuperar, usar o administrar un recurs d'informació».

La característica principal de les metadades és la seva naturalesa **referencial**, és a dir, es refereixen a «una altra cosa» que no són elles mateixes.

Hi ha una gran quantitat de metadades associades a qualsevol fet quotidià, per exemple, una trucada de telèfon, particularment les trucades des de dispositius mòbils. Les metadades més òbvies d'una trucada són els números de telèfon de la persona que truca i del destinatari. A continuació, hi ha l'hora i la durada de la trucada. I per a les trucades realitzades des de telèfons intel·ligents, hi ha les ubicacions geogràfiques de les persones involucrades (la majoria d'aquests dispositius tenen la funcionalitat GPS o pel mesurament de l'abast de les torres de telefonia mòbil en què es troben).

Així, es pot considerar que les metadades es comporten com una informació que **descriu un recurs**. De la mateixa forma que la informació nutricional d'un aliment envasat: les metadades són informació sobre el contingut.

Les metadades poden descriure una gran varietat d'informació, com:

- L'assumpte del recurs.
- Els creadors del recurs.
- La informació tècnica necessària per emmagatzemar i accedir al recurs.
- Els drets legals del recurs.

Les metadades es poden usar per buscar, reunir i mantenir recursos durant llargs períodes de temps. D'aquesta manera es considera tota la informació que un usuari necessitarà per saber si el recurs satisfà les seves necessitats o si pot usar-lo i com. Les metadades poden indicar de què tracta la informació, com usar-la, si necessita permisos i on obtenir aquesta informació. D'aquesta manera, les metadades permeten que les cerques siguin més significatives.

Metadada

Del grec *meta* 'més enllà' i del llatí *datum* 'allò donat'.

VALOR NUTRITIVO MEDIO (aprox. por 100 g. de producto)	
Valor Energético	338 kcal 1.437 kJ
Proteínas	7,9 g
Hidratos de Carbono	75,3 g
De los cuales azúcares	0,2 g
Grasas	0,5 g
De las cuales saturadas	0,1 g
Fibra Alimentaria	2,1 g
Sodio	0,0 g

Figura 1. Exemple de metadades: etiqueta nutricional d'un aliment envasat

Font: es.wikipedia.org/wiki/archivo:Etiqueta_nutricional_de_un_alimento_envasado.JPG

1.1. Diferència entre dades i metadades

Encara que puguin semblar similars, fins i tot en el nom, les dades i les metadades no són el mateix. L'explicació més senzilla és que les dades, en essència, són el producte d'informació real, mentre que les metadades són dades sobre aquestes dades.

En el context d'un arxiu de vídeo les dades són els fotogrames del vídeo que es visualitza. El nom de l'arxiu i el nombre de fotogrames, a més de la informació sobre els drets d'autor, no formen part d'aquesta presentació, sinó que descriuen les dades. Això són les metadades.

Les dades en si es denominen **dades essencials** perquè són la veritable essència de la informació que s'emmagatzema. Les dades essencials són les que tenen drets de propietat i valor comercial, mentre que les metadades ajuden a trobar, administrar i distribuir aquestes dades essencials.

1.2. Tipus i funcions

Els tipus de metadades depenen tant del context com del domini de l'aplicació. L'any 1998, el Getty Research Institute, en una publicació sobre metadades, va identificar cinc tipus de metadades i les seves funcions:

1) **Metadades administratives.** Metadades utilitzades per a la gestió i administració dels recursos d'informació. Per exemple, adquisició de la informació, drets i seguiment de la reproducció, requisits legals d'accés i informació d'ubicació.

2) **Metadades descriptives.** Metadades utilitzades per identificar i descriure recursos d'informació relacionats. Per exemple, catalogació, trobar ajudes o índexs especialitzats.

3) **Metadades de preservació.** Metadades relacionades amb la gestió de la conservació dels recursos d'informació o accions per protegir els recursos físics o digitals.

4) **Metadades tècniques.** Metadades relacionades amb el funcionament dels sistemes o comportament de les metadades. Per exemple, informació sobre maquinari i requisits de programari, formats, taxa de compressió o dades d'autenticació i seguretat.

5) **Metadades d'ús.** Metadades relacionades amb el nivell i tipus d'ús dels recursos d'informació. Per exemple, cerca i seguiment dels usuaris.

L'any 2004, l'Organització Nacional d'Estàndards d'Informació (NISO) d'EUA agrupa els tipus anteriors de metadades en dues categories i afegeix les metadades estructurals:

- 1) **Metadades descriptives**, que descriuen un recurs per a finalitats d'identificació i cerca.
- 2) **Metadades estructurals**, que indiquen com s'organitzen els recursos.
- 3) **Metadades administratives**, que proporcionen informació d'administració dels recursos, incloent les metadades tècniques, de gestió de drets i de preservació.

2. Estàndards de metadades

Les metadades estan formades per elements de dades associades al recurs (elements de metadades). Els **esquemes de metadades** o **models de metadades** són conjunts d'elements de metadades dissenyades per a un propòsit específic, com descriure un tipus particular de recurs. Quan un esquema de metadades està definit i és àmpliament acceptat, és adoptat per la majoria de la indústria. I quan una especificació és àmpliament reconeguda i adoptada, es converteix en un **estàndard de metadades**.

No hi ha un únic estàndard en metadades, sinó que hi ha nombrosos estàndards de metadades o especificacions de metadades que es poden adaptar per satisfer necessitats concretes del context.

2.1. Tipus i estructura de metadades

Els registres de metadades tenen tres components: sintaxi, estructura i semàntica.

La **sintaxi** són els estàndards de codificació que funcionen com un contenidor per a l'estructura i la semàntica, un conjunt de regles que permeten la interpretació dels continguts.

Com a exemples d'estàndards de codificació tenim:

- RDF (marc de descripció dels recursos)
- XML (llenguatge d'etiquetatge extensible)
- JSON-LD (JSON per a dades enllaçades)

L'**estructura** s'ajusta dins de la sintaxi i està composta per un esquema de metadades.

Alguns exemples d'estructures són:

- Dublin Core
- Categories per a la descripció d'obres d'art (CDWA Lite)
- Descripció d'arxiu codificat (EAD)

Mentre que l'estàndard de codificació indica a un sistema informàtic com interpretar un registre de metadades, l'estructura indica a un usuari humà com interpretar la informació continguda en etiquetes individuals (per exemple, creador, títol, tema, etc.). Els estàndards d'estructura es poden considerar com a conjunts d'etiquetes (*tags*) preparats.

La peça final d'un registre de metadades és la **semàntica** o els estàndards de contingut. Els estàndards de contingut defineixen com ha de ser la informació de dins de les etiquetes o *tags*.

L'ús d'estàndards de contingut garanteix la uniformitat de la informació continguda en les etiquetes, és a dir, és una peça fonamental per produir metadades interoperables.

2.2. Dublin Core Metadata Initiative

Dublin Core Metadata Initiative (DCMI) és una organització internacional amb seu a Dublín (Ohio, EUA) que es finança per mitjà de projectes i subvencions i en què el treball es realitza per voluntaris de tot el món.

Els objectius d'aquesta iniciativa són:

- Desenvolupar estàndards de metadades per a la recuperació d'informació a internet entre diversos dominis d'informació.
- Definir marcs de treball per a la interoperabilitat entre conjunts de metadades.
- Facilitar conjunts de metadades especialitzades per a una comunitat o disciplina.

De forma general, les metadades Dublin Core consten de:

- 15 elements bàsics (vegeu taula 1).
- Qualificadors per a detalls addicionals.
- Tipus de dades (formats de data/temps, tipus MIME).

Taula 1. Conjunt d'elements de Dublin Core

Etiqueta	Descripció
DC. Title	Títol: nom donat a un recurs.
DC. Creator	Autor: entitat primàriament responsable de la creació del contingut intel·lectual del recurs.
DC. Subject	Matèries i paraules clau: tema del contingut del recurs.
DC. Description	Descripció: descripció del contingut del recurs.

Etiqueta	Descripció
DC. Publisher	Editor: entitat responsable de fer que el recurs estigui disponible.
DC. Contributor	Col·laborador: entitat responsable de fer contribucions al contingut del recurs.
DC. Date	Data: data associada amb un esdeveniment en el cicle de vida del recurs.
DC. Type	Tipus: naturalesa o categoria del contingut del recurs.
DC. Format	Format: manifestació física o digital del recurs.
DC. Identifier	Identificació: referència no ambigua per al recurs dins d'un context donat.
DC. Source	Font: referència a un recurs del qual es deriva el recurs actual.
DC. Language	Idioma: idioma del contingut intel·lectual del recurs.
DC. Relation	Relació: referència a un recurs relacionat.
DC. Coverage	Cobertura: extensió o àmbit del contingut del recurs.
DC. Rights	Drets: informació sobre els drets de propietat i sobre el recurs.

Dublin Core, per tant, es pot considerar com un petit llenguatge per realitzar una classe particular de declaracions sobre recursos. En aquest llenguatge hi ha dues classes de termes: **elements** (noms) i **qualificadors** (adjectius).

2.3. IEEE/LTSC LOM i IMS Global Learning Meta-Data

Es tracta de l'estàndard per excel·lència per a metadades d'objectes educatius i està patrocinat pel Comitè d'Estandardització de Tecnologies Educatives del IEEE (Institute of Electrical and Electronics Engineers).

LOM (*Learning Object Metadata*) va ser aprovat com l'estàndard IEEE 1484.12.1:2002 sobre metadades per a objectes d'aprenentatge com un estàndard obert per a la descripció dels «objectes d'aprenentatge».

LOM defineix i especifica un esquema de metadades amb atributs, les seves definicions i una estructura jeràrquica que les relaciona. No inclou informació sobre com representar aquestes metadades o amb quins mecanismes es pot transmetre i processar aquesta informació. Una iniciativa d'IMS Global Learning Consortium ha especificat models per a l'ús de les metadades i una representació en XML que proporcionen els mecanismes per ser usats a la pràctica.

El principal objectiu d'aquest estàndard és facilitar la cerca, l'avaluació, l'adquisició i l'ús de recursos docents.

Així, els estàndards LOM se centren en el conjunt mínim de propietats que permeten que els objectes educatius siguin gestionats, situats i avaluats.

Les metadades LOM s'organitzen en una estructura en arbre amb un node arrel que correspon al node que s'està descrivint (LOM). Depenent de l'arrel, tenim subelements que poden contenir altres subelements. Per a cada element de la jerarquia s'indica la definició, el tipus de dades, els valors permesos i la multiplicitat.

L'Esquema de base de LOM es compon de 9 categories i 47 elements (vegeu taula 2).

Taula 2. Esquema de base de LOM

Categories	Elements
1. General	Identificador Títol Entrada de catàleg Idioma Descripció Descriptor Cobertura Estructura Nivell d'agregació
2. Cicle de vida	Versió Estatus Altres col·laboradors
3. Meta-metainformació	Identificador Entrada de catàleg Altres col·laboradors Esquema de metadades Idioma
4. Tècnica	Format Grandària Ubicació Requisits Comentaris sobre la instal·lació Altres requisits per a plataformes Durada
5. Ús educatiu	Tipus d'interactivitat Tipus de recurs d'aprenentatge Nivell d'interactivitat Densitat semàntica Usuari principal Context: àmbit educatiu Edat Dificultat Temps previst d'aprenentatge Descripció Idioma
6. Drets	Cost Copyright i altres restriccions Descripció

Categories	Elements
7. Relació (amb altres recursos)	Tipus: naturalesa de la relació amb el recurs principal Recurs: recurs principal a què es refereix aquesta relació
8. Observacions	Persona Data Descripció
9. Classificació	Finalitat Nivell taxonòmic Descripció Descriptor

2.4. Més sobre estàndards

A continuació, es presenten algunes organitzacions que proporcionen estàndards de metadades sobre temes especialitzats:

- **Creative Commons**, que conté llicències amb les quals es poden usar metadades sobre drets de propietat i atorgar drets.
- **Biblioteca del Congrés dels Estats Units d'Amèrica** segueix sent la font de referència per a la catalogació en biblioteques i manté la Classificació de la Biblioteca del Congrés (LCC)¹ i els Epígrafs Temàtics de la Biblioteca del Congrés (LCSH).²
- **FGCD** (Federal Geographic Data Committee) van ser pioners en la implantació de les metadades. Les metadades són especialment importants per a les dades geoespacionals amb els quals generar mapes fiables.
- **Ecological Metadata Language** (EML) centrat en dades sobre ecologia.
- **Water Markup Language** (WaterML) amb especial interès en sèries temporals sobre observacions hidrològiques.

⁽¹⁾De l'anglès, *Library of Congress Classification*.

⁽²⁾De l'anglès, *Library of Congress Subject Headings*.

Bibliografia

Alemu, G.; Stevens, B. (2015). *An Emergent Theory of Digital Library Metadata*. Cambridge: Chandos Publishing.

Gidley, S.; Castanedo, F. (2017). *Understanding Metadata*. O'Reilly Media.

Lei Zeng, M.; Qin, J. (2016). *Metadata* (2a. ed.). Chicago: American Library Association.

Lubas, R.; Schneider, I.; Jackson, A. (2013). *The Metadata Manual*. Cambridge: Chandos Publishing.

Pomerantz, J. (2015). *Metadata*. Massachusetts: The MIT Press.

Wootton, C. (2009). *Developing Quality Metadata*. Massachusetts: Focal Press.

